空间 VLBI 数据记录格式 RDF 解析与数据解码*

张浩 [1,2] 张娟 [1,4] 刘磊 [1,4] 郑为民 [1,3,4#] 芮萍 [1]

（1.中国科学院上海天文台，上海 200030； 2.中国科学院大学，北京 100049； 3.中国科学院射电天文重点实验室，江苏 南京 210008； 4.上海市导航定位重点实验室，上海 200030）

摘 要：空间 VLBI 由于基线更长，在相同观测频率下，可获得比地基 VLBI 更高的分辨率。我国正在推进空间 VLBI 科学与技术研究。俄罗斯 RadioAstron 空间 VLBI 项目于 2018 年结束，其成功经验可资借鉴。该项目采用不同于常规地面 VLBI 的专用数据格式 RDF（RadioAstron Data Format）记录观测数据，利用配备的星载或地面上行氢原子频标提供频率基准，数据经采集量化后传输至地面站，打上时标并记录于硬盘。为处理 RadioAstron 数据，本文分析了 RDF 格式，完成了 RDF 数据解码及 VLBI 相关处理，然后对比分析了 Mark5B、VDIF 等地面 VLBI 通用数据格式和 RDF 格式的特点。这些工作将为我国未来空间 VLBI 数据格式设计与处理积累经验。

关键词：空间 VLBI；RDF 格式；数据解码；Mark5B

0 引言

甚长基线干涉测量技术(VLBI)是目前角分辨率最高的天文观测技术，在天文观测、大地测量和深空探测等领域得到了广泛应用[1]。其分辨率取决于天线的口径和两天线之间的基线长度。随 VLBI 科学研究与火星探测等工程任务的推进，对 VLBI 空间分辨率提出了越来越高的要求。为进一步提高 VLBI 网的空间分辨率，需要增加基线长度。受到地球尺度的限制，地面 VLBI 基线长度无法超过地球直径，因此，对空间 VLBI（Space Very Long Baseline Interferometry, SVLBI）技术的研究变得越来越迫切。

早在 20 世纪 80 年代初，欧洲空间局(ESA)和 NASA 联合向 ESA 提出 QUASAT(QUASAR SATELLITE)计划的立项建议，计划是在太空放置一个 15m 口径的天线，并在预研阶段就 SVLBI 的技术要求、观测性能、数据分析成图能力及预期科学目标等做了广泛研究，为后来的 SVLBI 发展提供了经验[2]。随后的 1986-1988 年间，NASA 利用 TDRSS(Tracking Data and Relay Satellite System)测控中继卫星上的 4.9m 天线和地面上的两个 64m 射电望远镜对一批致密河外射电源在 2.3GHz 和 15GHz 频率上进行了三次 VLBI 观测，大部分源检测到干涉条纹，证实了 SVLBI 技术的可行性[2]。在此期间，前苏联提出 RadioAstron 计划，受到苏联解体的影响，该项目曾中断执行，后于 2010 年恢复并正式进行 SVLBI 发射及观测任务。日本也于 1997 年 2 月提出 HACLA（Highly Advanced Laboratory for Communications and Astronomy)空间干涉测量计划，首次提供了空间 VLBI 多望远镜成图观测条件，旨在实现活动星系核和射电星的高分辨率成图[2]。

目前只有日本和俄罗斯完成了正式的空间 VLBI 项目。上海天文台正在推进我国空间 VLBI 发展，

* 本文受到中科院空间科学预研项目（XDA15016100）、国家自然科学基金（11973011，11573057，11803070）、国家科技基础条件平台项目“国家基础科学数据共享服务平台”（DKA2017-12-02-09）、上海市空间导航与定位技术重点实验室（15DZ2271700，ZZXT-201902）、中科院关键技术人才、上海市领军人才、国家万人计划、探月工程等多个项目支持。

通讯作者：郑为民，zhwm@shao.ac.cn

其中“空间低频射电天文台”项目，计划发射两个大口径低频射电望远镜至近地空间，形成地—空、空—空干涉能力。地月 VLBI 计划也在推进中。由于空间与地面环境的巨大差异性，空间 VLBI 研究存在一些亟待解决的问题，比如，怎样保持空间射电望远镜和地面跟踪站之间稳定的信号链接，如何建立实用有效的空间 VLBI 时延模型等。本文主要讨论空间 VLBI 数据终端记录格式的问题，通过对 RadioAstron 原始观测数据开展研究，解析其数据记录格式，完成数据解码，使之能够与地面 VLBI 观测数据进行相关处理。现阶段，国际上使用的地基 VLBI 格式有很多种，而空间 VLBI 格式（如 RDF）只有日本和俄罗斯在使用。

MARK5B 数据格式是美国麻省理工学院 Haystack 天文台设计的 VLBI 数据终端记录格式，采用硬盘存储数据，存储成本低，便于通过互联网传输数据，并促进了 e-VLBI 技术的发展。MARK5B 格式严格遵守 VSI-H(The VLBI Standard Interface Hardware Interface Specification)规范，能够兼容所有采用 VSI 接口的数据记录格式，有效降低了 VLBI 数据记录硬件设施更新的频率[3]。MARK5B 格式数据分为多个磁盘帧，每帧由固定大小的帧头和紧随其后的数据块组成，帧头包含同步字、CRC 校验码、秒内帧号和起始 UTC 时间等信息，能够有效保证数据的安全性[4]。由于存在丢帧、乱帧的现象，为监测数据的质量，MARK5B 需要每秒统计一次数据有效率，作为数据可靠性和网络诊断的参考[4]。中国 CVN 网的 5 个台站均使用上海天文台自行研制的中国数据获取系统（China Data Acquirement System, CDAS）[5]。CDAS 采用 MARK5B 格式记录原始观测数据，已成功完成了中国探月工程任务。

VDIF(VLBI Data Interchange Format)是一种新的 VLBI 标准化数据交换格式，支持实时和准实时 e-VLBI 以及磁盘文件存储[6]。VDIF 格式基于自适应长度的数据帧制定，帧头携带有帧起始时间、帧长和线程号等关键信息，数据帧长度由帧头中的帧长指定，用户可以自由设定数据帧长度，以最佳地匹配所选择的数据传输协议[6]。例如，在实时网络传输情况下，适当选择数据帧的长度使得每个线上数据包只携带一个数据帧。另外，VDIF 还支持多线程。同一频率通道的时间序列数据帧存储在同一线程，不同频率通道的数据存储在多个并行的数据线程。对于通过串行网络的数据传输，多个并行的数据线程合并为单个串行的“数据流”[6]。很明显，VDIF 的优势在于数据网络交换和数据兼容性，但目前支持 VDIF 格式的设备为数不多，应用受到限制。

俄罗斯空间 VLBI 项目 RadioAstron 采用 RDF 格式用于地面数据采集。由于 RadioAstron 没有搭载大容量存储设备，采集的空间 VLBI 数据必须实时发送至地面。实际上，空间射电望远镜数据采集包含 VLBI 空间天线和地面跟踪站两部分。VLBI 空间天线负责接收原始 VLBI 观测信号；地面跟踪站负责锁定空间天线的下行信号，实时采集空间天线发送的 VLBI 数据并以 RDF 格式记录。

VSOP 和 RadioAstron 的空间 VLBI 射电望远镜系统，采用了不同于常规地面 VLBI 站的结构（图 1）。为实现对空间天线椭圆轨道的全天候实时跟踪，必须在全球建立多个地面跟踪站。地面跟踪站的主要作用之一是为空间望远镜提供一个稳定的频率参考[7]，确保精准控制空间天线的采样率。此外，由于空间天线时间频率系统的不稳定性，为了确保数据接收时刻的准确性，地面跟踪站记录空间 VLBI 数据到达该跟踪站的时刻。该时刻随着数据字段的增长而增加，相关处理时由处理机根据计算的光行时（光行时等于空间天线与地面跟踪站的距离除以光速）反算出每个数据字段在空间天线上对应的时刻。出于满足科学需求基础上数据格式最简化的考虑，一个 RDF 文件只需要在起始的文件头打上时间标签，后续的数据接收时刻根据数据采样率依次推算即可。空地数据传输不支持复杂的数据格式，因此 RDF 结构简单，在未来的空间 VLBI 项目中或将继续使用。

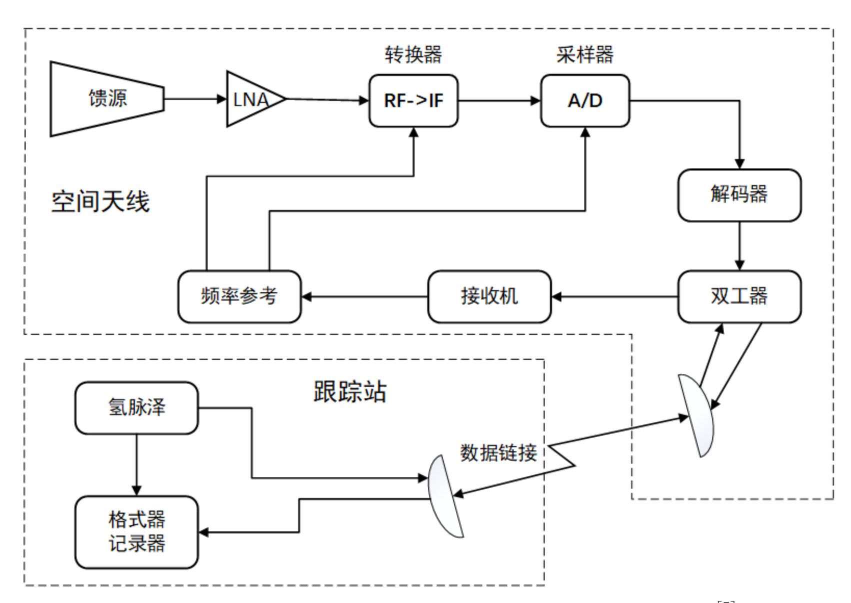


图 1　VLBI 空间射电望远镜系统组成结构示意图[7]

1 RDF 数据格式

RDF 格式是由空间射电望远镜 RadioAstron 采集 VLBI 数据并以特定频率实时发送至地面，由地面跟踪站记录并输出的一种基于磁盘存储的 VLBI 数据格式。

地面跟踪站会在一组文件中记录时间连续的射电天文数据，除了最后一个文件外，每个文件的数据长度约为 15 分钟的数据量。一个文件记录到指定的数据量后，系统便会新建一个文件记录其后的数据，直至达到指定数据量后再建立新的文件……这样周而复始，直到数据记录终止。

文件命名按照“测站名+UTC”的形式，以文件记录的第一个 VLBI 数据对应的 UTC 时间为准，这个时间一定是整数秒。具体形式如下：

“station_YYYYDDDHHMMSS.rdf”

其中，“station”为测站名，“YYYY”为公历年号，“DDD”为观测当天的年积日，“HH”、“MM”和“SS”分别为 UTC 时刻对应的时、分、秒。

RDF 的文件结构不像 MARK5B 那样分成若干个等长的磁盘帧，每一帧用固定格式的帧头作为标识，帧头包含该帧的相关信息。与之不同的是，单个 RDF 格式文件只包含一个文件头，其后的所有数据都是 VLBI 原始观测数据，如图 2 所示。

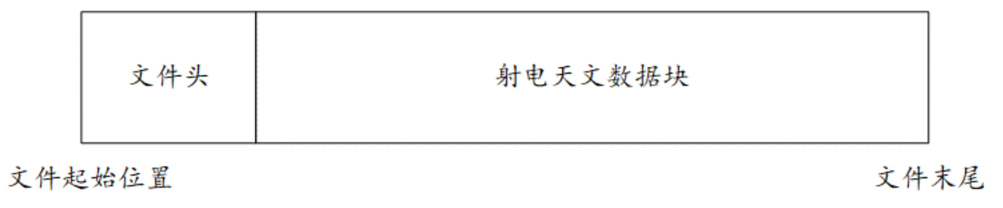


图 2 RDF 文件结构示意图

1.1 文件头格式

RDF 文件头固定占用 256 字节磁盘空间，VLBI 原始数据从第 257 字节开始不间断记录，直到停止数据采样。文件头记录了数据起始 UTC 时间、数据传输速率和接收模式等重要信息。其结构如下：

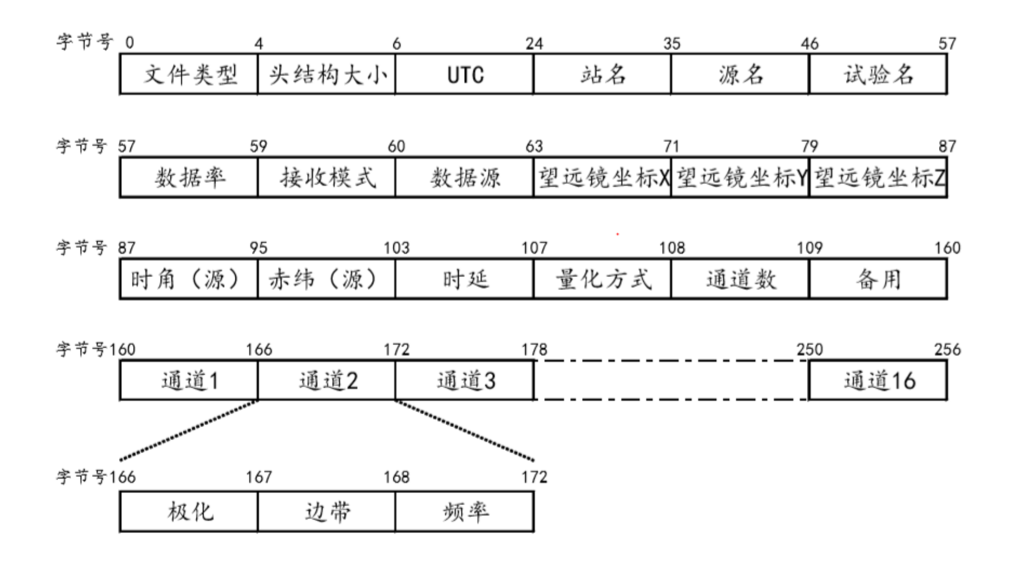


图 3 RDF 文件头格式

图 3 标记了头文件每个数据项的起始字节号和占用空间的大小。文件类型是一个由 4 个 ASCII 字符组成的同步字，用来识别射电天文数据文件，这里必须为“RDF1”。头文件大小一定是 256 字节，用二进制数“00000001 00000000”表示。UTC 表示数据开始记录的日期和时间，以“YYYY DDD-hh:mm:ss”的形式存储为 ASCII 字符串，固定占用 18 字节空间。数据率即数据传输速率，按 2 字节的 ASCII 字符串存储，RadioAstron 的数据率为 16Mb/s，故存为“16”。接收模式用一个字符表示，有“I”、“Q”、“A”、“R”四种，RadioAstron 采用“R”模式。数据源存储为 3 字节字符串，地面望远镜记作“DAS”，空间望远镜记作“DEC”。字节 63-160 由相关处理机填入，相关处理机从 VEX 文件中获取这些参数，其中望远镜坐标由轨道数据计算得到。字节 161-256 划分为 16 个通道，每个通道占用 6 字节空间，记录该通道的频率、边带和极化。

1.2 数据字格式

数据采集终端从接收机送来的中频信号 IF 中选出所需要观测的频率通道，然后经过频率变换至基频信号，再经过数字化采样和编码后记录在磁盘上[8]。实际使用中常常需要同时选出多个频率通道[8]。

数据块中数据流的记录格式是由通道数和量化方式决定的，每个通道对应一条比特流。MARK5B 硬盘记录器能够记录的比特流数量只能是 1、2、4、8、16、32 这 6 种[9]，图 4 列出了通道数为 8 或 16、量化方式为 1bit 或 2bit 时的数据单元记录格式。

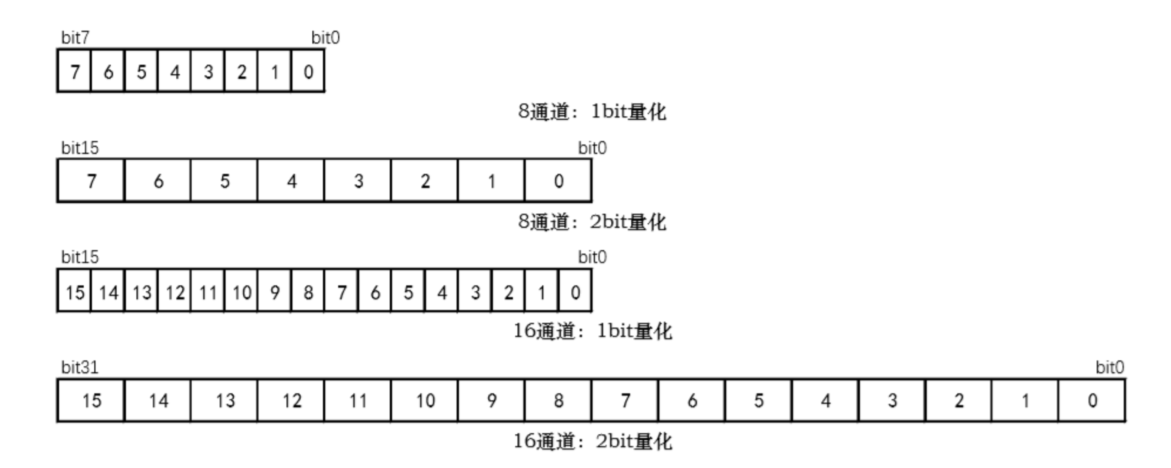


图 4 数据块中数据单元的格式

以 16 通道、1bit 量化为例，数据块部分的第 0、16、32、48、64、80……bit 的数据是按照先后顺序依次记录的通道 0 中的比特流，第 1、17、33、49、65、81……bit 的数据是通道 1 中的比特流，依此类推可以得到其他通道的比特流。图 4 只是一个示意图，实际上各通道并不是如此规律地排布，多 bit 量化模式下通道内的 bit 位也不一定是相邻的。通道的数目以及数据单元中的 bit 位的排布由地面跟踪站给出的 VEX 文件指定。

图 5 为 VEX 文件中跟踪站指定 bit 位排布的字段样例。图 6 给出了 4 通道 2bit 量化下 bit 位排布的示意图。

```
def MARK5B.4Ch2bit1to1;
* mode =  1    stations =Wb
*    format = MARK5B, and fan-out = 1
*    mode requires  32.00Mb/s/tr; stations using disks
     track_frame_format = MARK5B;
     fanout_def =   : &CH01 : sign : 1:  4;
     fanout_def =   : &CH01 :  mag : 1:  7;
     fanout_def =   : &CH02 : sign : 1:  6;
     fanout_def =   : &CH02 :  mag : 1:  2;
     fanout_def =   : &CH03 : sign : 1:  5;
     fanout_def =   : &CH03 :  mag : 1:  3;
     fanout_def =   : &CH04 : sign : 1:  8;
     fanout_def =   : &CH04 :  mag : 1:  9;
```

图 5 VEX 文件中的 bit 位排布规则

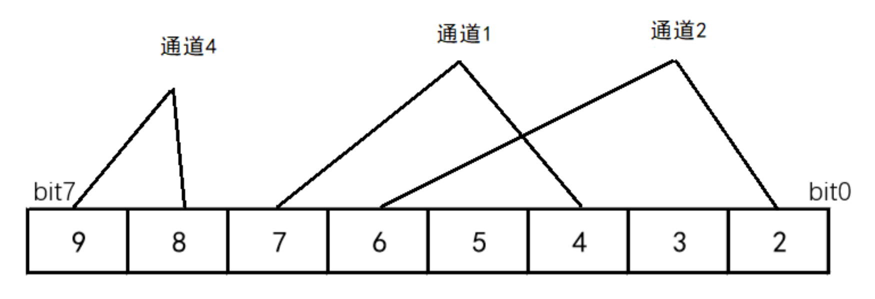


图 6 数据单元 bit 位排布示意图

2 VLBI 数据解码

数据解码是进行相关处理之前的数据预处理过程，对于不同的数据格式，采用的解码算法也不同。设计解码算法的关键在于了解数据存储结构。我们通过国际空间 VLBI 合作项目获取了部分 RDF 数据源文件，研究并针对 RDF 设计了实用可行的解码算法。

解码算法具体流程如图 7 所示。由于 RDF 只有一个文件头，首先我们读取文件头，提取出数据解码最为关键的两个信息：起始 UTC 时间和数据率，再根据数据率计算每秒的数据量 T,然后从数据块中循环读取数据量大小为 T 的数据到新建的以整秒 UTC 时间命名的文件当中，直至数据读取完毕。

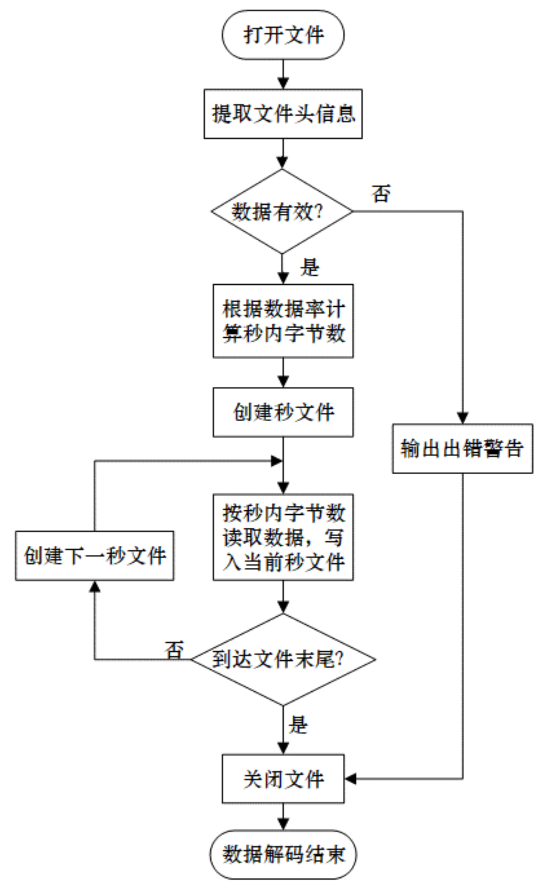


图 7 RDF 文件数据解码流程图

3 数据验证

通过对解码后数据的相关处理，可以验证上述解码过程的正确性。我们采用的 RDF 数据来自 2014 年俄罗斯空间 VLBI 网对河外射电源 0823+033 的一次观测试验。参与此次观测的望远镜除了 RadioAstron 之外，还有美国的 Arecibo 和荷兰的 Westerbork 两个地面射电望远镜，其中 Arecibo 数据记录格式为 Mark5A，Westerbork 数据记录格式为 Mark5B。在上海天文台 VLBI 中心进行相关处理之前，每个观测台站的数据都需经过解码，转换为特定格式的预处理输入数据。经自研的软件相关处理机 SCORR 处理后，成功获得了清晰的干涉条纹（图 8）。

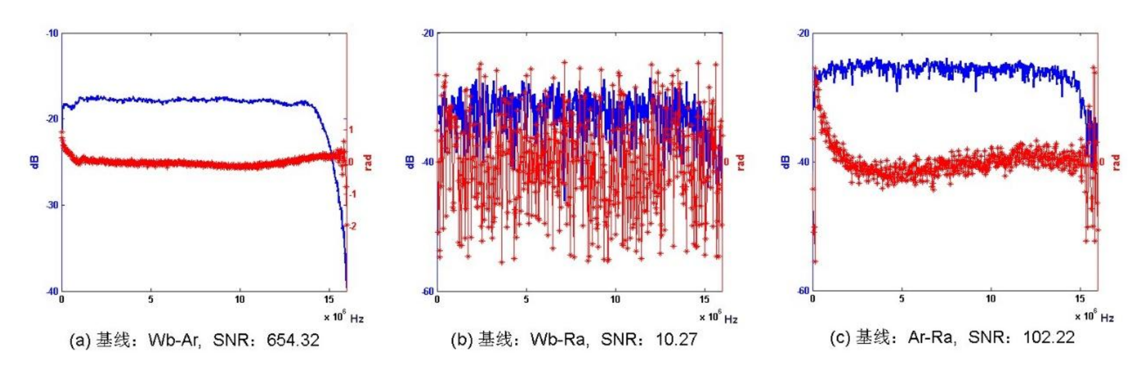


图 8 空间 VLBI 数据相关处理干涉条纹

图 8 给出了空间 VLBI 数据相关处理结果，其中红色部分表示相位谱，蓝色部分表示幅度谱。Ar-Ra 基线存在明显的干涉条纹（Ar 表示 Arecibo，Ra 表示 RadioAstron）。结果说明参与相关处理的经解码得到的预处理数据是有效的，验证了 RDF 格式数据解码的正确性。Wb-Ra 基线信噪比较低，可能的原因是两望远镜口径相对较小（Westerbork 是由 14 个 25 米网状天线组成的综合孔径射电望远镜，等效口径为 66 米，综合灵敏度没有 Arecibo 高，空间射电望远镜 RadioAstron 口径 10m），加之基线很长（约 20 万千米），导致条纹较弱。Arecibo 口径相对较大（300 米），条纹较清晰。

我们的信噪比计算方法与 HOPS[1]（Haystack Observatory Postprocessing System）一致（见式(1)），Wb-Ar 基线条纹的信噪比为 654.32，Wb-Ra 基线条纹的信噪比为 10.27，Ar-Ra 基线条纹的信噪比为 102.22。

$$SNR = Amp\frac{\sqrt{2BT_{ap}/n_{ap}}}{n_{lags}-1} \quad (1)$$

其中，Amp 表示所有积分周期和通道内所有频点相加后求得的幅值（去除第一个频点的直流分量），B 为通道带宽，T_{ap} 为单位积分周期，n_{lags} 为每通道的频率点数，n_{ap} 为所有积分周期和通道的权值总和。权值从相关处理机的输出文件中读取。

4 三种 VLBI 数据格式对比

表 1 列出了常用地面和空间 VLBI 所使用的 VDIF、MARK5B 和 RDF 三种数据格式的简要特点。

RDF 格式由于只有一个文件头的额外数据记录，其余的记录都是原始 VLBI 数据 ，相较而言 RDF 具有更高的记录效率。为保证数据记录的可靠性，RadioAstron 项目采用了一个额外的解码器，当解码器检测到数据丢失或者发生偏移，解码器就会新建一个 RDF 文件开始记录。由于数据丢失或偏移出现的频率比较低，且 RDF 文件避免了其他两种格式中帧头数据的记录，这种方法同时兼备了较好的数据记录可靠性和记录效率。

MARK5B 格式将数据文件划分成多个磁盘帧，每个磁盘帧帧长固定为 10000 字节，由额外的 16 字节帧头引导。虽然不可避免地会出现丢帧、帧号乱序和帧长不足等情况，但是通过帧头中的 UTC 时间和秒内帧号可以确定当前帧在时间序列中的位置，并采用相应的数据纠错方法可以提高整个数据文件的数据有效率[4]，因此 MARK5B 格式具有较高的数据可靠性。此外，MARK5B 格式是完全按照 VSI 标准制定的，与现有的 VLBI 系统保持广泛的兼容性[10]。目前国际上使用最多的 VLBI 终端系统是 MK 系列终端系统[11]，因此 MARK5B 格式具有应用的广泛性。

VDIF 格式同样采用多个数据帧的形式保存数据，与 MARK5B 不同的是，VDIF 采取可变帧长（可变帧长指的是针对不同的传输协议可以设定与其对应的帧长以适应单个数据包的容量，同一个观测

1 参考 Haystack 官方网站：https://www.haystack.mit.edu/tech/vlbi/hops.html

文件数据帧的长度是固定的）的机制以确保所采用的传输协议的数据包刚好容纳一个 VDIF 数据帧[6]，这样在传输过程中如果发生丢包事件或者数据包未能按序交付，接收端也能够通过数据包内的帧头信息恢复数据时序以及补足缺失数据（用 0 补足）。如果将 MARK5B 等传统 VLBI 数据记录格式用于网络传输，可能会因为丢包使得数据位发生移位，而 VLBI 相关处理对数据位移位非常敏感[4]，由此导致接收端接收的数据文件不可靠。相比之下，VDIF 格式的网络传输安全性更高。VDIF 格式设计的目的是为了解决当前国际上数据终端记录格式多样化导致数据交换不便的问题，除了方便数据交换之外，VDIF 格式也支持磁盘存储[6]。随着数据终端设备的更新换代，VDIF 将会得到更加广泛的应用，成为兼具数据记录和数据网络交换功能的宽带 VLBI 通用数据格式。

表1 MARK5B、RDF和VDIF格式对比

	MARK5B	RDF	VDIF
说明	地基VLBI数据终端记录格式	空间VLBI数据终端记录格式	VLBI数据交换格式
用途	地面天线观测数据记录	空间VLBI观测数据记录	e-VLBI数据网络传输、实时e-VLBI和传统VLBI数据间的格式转换
帧头	有多个帧头，占用4X32个比特位，共16字节，包含同步字、秒内帧号、时间码、CRC校验码等信息	只有一个帧头，占用256字节，包含采样数据的起始UTC、测站信息、传输速率等	有多个帧头，占用32字节，包含当前帧距参考历元的时间秒数、秒内帧号、帧长等
数据块	原始VLBI数据，占用1万字节，每个数据块紧跟对应的帧头	仅有一个数据块，帧头后面全是原始VLBI数据	原始VLBI数据，每个数据块紧跟对应的帧头，数据块容量由帧长决定，帧长包含帧头占用空间
其他			支持多线程

综上所述，RDF 是一种应用于空间 VLBI 地面跟踪站的数据记录格式，MARK5B 格式是现阶段应用范围最广的地基 VLBI 数据记录格式，而 VDIF 以其灵活的设计和较高的网络传输安全性可能成为未来一段时间的通用数据格式。需要说明的是，MARK5B 格式由 Haystack 天文台于 2005 年前后设计完成[3]，同样基于磁盘的 VLBI 数据记录格式 MARK5A 于 2002 年设计完成[10]，而 RDF 格式于 RadioAstron 预研阶段（20 世纪 80 年代）研制成功，早于 MARK5B 和 VDIF 格式。随着时间的推移，未来 RDF 也可能被更加先进的数据记录格式所取代。

5 总结

本文在国内首次分析了俄罗斯空间射电望远镜 RadioAstron 特有的数据记录格式 RDF，完成了 RDF 格式的实测空间 VLBI 数据解码，在此基础上利用 SCORR 软件处理机成功进行了空间 VLBI 相关处理及验证，掌握了 RDF 格式解码方法。最后对比了 RDF 格式及 RDF 相关现有地基 VLBI 格式，分析了它们各自具有的特点。这些研究工作有助于后续深入分析处理更多的 RadioAstron 实测数据，推进我国空间 VLBI 数据处理技术的研究。

致谢

在研究过程中，俄罗斯科学院列别捷夫物理所天文宇航中心 Andrianov Andrey 博士对 RDF 格式的一些细节问题在邮件中给予了热心回复；上海天文台国际访问学者 Tetsuro Kondo 教授和童锋贤博士对数据相关处理工作给予了大力支持，向他们表示感谢。

参考文献:

[1] 朱新颖,李春来,张洪波.深空探测 VLBI 技术综述及我国的现状和发展[J].宇航学报,2010,31(8): 1894-1897.
[2] 沈志强.空间 VLBI 的研究现状[J].天文学进展,1998(2):117-134.
[3] MIT Haystack Observatory. Mark5B System User's Manual[EB/OL].https://www.haystack.mit.edu/tech/vlbi/mark5/docs/Mark%205B%20users%20manual.pdf, 2006.
[4] 郭迎,郑为民.一种高性能 VLBI 数据预处理方法及其软件实现[J].中国科学院上海天文台年刊,2013:114-122.
[5] 杨文军,郝龙飞.VLBI 终端系统的发展历史和未来展望[J].天文研究与技术,2012,09(4):374-381.
[6] Whitney A , Kettenis M , Phillips C , et al. VLBI Data Interchange Format (VDIF)[J]. European Vlbi for Geodesy & Astrometry Working Meeting, 2010, 194(2-5):156-160.
[7] Cornwell T E , Fomalont E B . Synthesis Imaging in Radio Astronomy II[J]. 1989.
[8] 钱志瀚.甚长基线干涉测量技术在深空探测中的应用[C].中国宇航学会深空探测技术专业委员会第七届学术年会论文集.2010.
[9] 朱新颖,张洪波,李春来.VLBI MK5B 通道检测软件的设计与实现[J].天文研究与技术,2011,08(3):236-241.
[10] 韦文仁,薛祝和.基于磁盘的新型 VLBI 终端系统——MK5A 终端系统[J].天文学进展,2004,22(3):269-274.
[11] 陈中,郑为民,陈肖.中国 e-VLBI 网的建立及应用[J].中国科学院上海天文台年刊, 2015,36:136-147.

A Study of Space VLBI Data Recording Format—RDF and its Data Decoding

Hao Zhang[1,2], Juan Zhang[1,4], Lei Liu[1], Weimin Zheng[1,3,4#], Ping Rui[1]

(1. Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, China; 4. Shanghai Key Laboratory of Navigation and Positioning, Shanghai 200030, China)

Abstract: Compared with ground VLBI, space VLBI provides higher angular resolution since its baseline is much longer. Our country has started the development of our own space VLBI science and technology. Russian space-VLBI project RadioAstron was finished in 2018. We can learn a lot from its successful experience. This project, which is quite different from ground-VLBI, uses a dedicated data format RDF (RadioAstron Data Format). Besides that, the frequency reference is provided by the on-board or ground-based uplinking hydrogen atomic frequency standard. The data is quantized and transferred to the ground station, and is then labeled with time-tag and recorded in the hard disk. In this paper, we analyze the RDF format, carry out RDF data decoding and VLBI correlation. Moreover, we compare RDF with other commonly used VLBI data format, e.g., Mark5B and VDIF. Our work will pave the way for the format designing and data processing of space-VLBI data in China.

Key words: Space VLBI; RDF format; Data decoding; Mark5B